



Computational Decomposition of Style for Controllable and Enhanced Style Transfer

Minchao Li, Shikui Tu^(✉), and Lei Xu^(✉)

Department of Computer Science and Engineering,
Center for Cognitive Machines and Computational Health (CMaCH),
Shanghai Jiao Tong University, Shanghai, China
{marshal.lmc,tushikui,leixu}@sjtu.edu.cn

Abstract. Neural style transfer has been demonstrated to be powerful in creating artistic images with help of Convolutional Neural Networks (CNN), but continuously controllable transfer is still a challenging task. This paper provides a computational decomposition of the style into basic factors, which aim to be factorized, interpretable representations of the artistic styles. We propose to decompose the style by not only spectrum based methods including Fast Fourier Transform and Discrete Cosine Transform, but also latent variable models such as Principal Component Analysis, Independent Component Analysis, and so on. Such decomposition induces various ways of controlling the style factors to generate enhanced, diversified styled images. We mix or intervene the style basis from more than one styles so that compound style or new style could be generated to produce styled images. To implement our method, we derive a simple, effective computational module, which can be embedded into state-of-the-art style transfer algorithms. Experiments demonstrate the effectiveness of our method on not only painting style transfer but also other possible applications such as picture-to-sketch problems.

Keywords: Style transfer · Representation learning · Deep neural network

1 Introduction

Painting art has attracted people for many years and is one of the most popular art forms for creative expression of the conceptual intention of the practitioner. Since 1990's, researches have been made by computer scientists on the artistic work, in order to understand art from the perspective of computer or to turn a camera photo into an artistic image automatically. One early attempt is Non-photorealistic rendering (NPR) [17], an area of computer graphics, which focuses on enabling artistic styles such as oil painting and drawing for digital images. However, NPR is limited to images with simple profiles and is hard to generalize to produce styled images for arbitrary artistic styles.

One significant advancement was made by [8], called neural style transfer, which could separate the representations of the image content and style learned

by deep Convolutional Neural Networks (CNN) and then recombine the image content from one and the image style from another to obtain styled images. During this neural style transfer process, fantastic stylized images were produced with the appearance similar to a given real artistic work, such as Vincent Van Gogh’s “The Starry Night”. The success of the style transfer indicates that artistic styles are computable and are able to be migrated from one image to another. Thus, we could learn to draw like some artists apparently without being trained for years.

Following [8], a lot of efforts have been made to improve or extend the neural style transfer algorithm. The content-aware style-transfer configuration was considered in [23]. In [14], CNN was discriminatively trained and then combined with the classical Markov Random Field based texture synthesis for better mesostructure preservation in synthesized images. Semantic annotations were introduced in [1] to achieve semantic transfer. To improve efficiency, a fast neural style transfer method was introduced in [13, 21], which is a feed-forward network to deal with a large set of images per training. Results were further improved by an adversarial training network [15]. For a systematic review on neural style transfer, please refer to [12].

The recent progress on style transfer relied on the separable representation learned by deep CNN, in which the layers of convolutional filters automatically learns low-level or abstract representations in a more expressive feature space than the raw pixel-based images. However, it is still challenging to use CNN representations for style transfer due to their uncontrollable behavior as a black-box, and thus it is still difficult to select appropriate composition of styles (e.g. textures, colors, strokes) from images due to the risk of incorporation of unpredictable or incorrect patterns. In this paper, we propose computational analysis of the artistic styles and decompose them into basis elements that are easy to be selected and combined to obtain enhanced and controllable style transfer. Specifically, we propose two types of decomposition methods, i.e., spectrum based methods featured by Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), and latent variable models such as Principal Component Analysis (PCA), Independent Component Analysis (ICA). Then, we suggest methods of combination of styles by intervention and mixing. The computational decomposition of styles could be embedded as a module to state-of-the-art neural transfer algorithms. Experiments demonstrate the effectiveness of style decomposition in style transfer. We also demonstrate that controlling the style bases enables us to transfer the Chinese landscape paintings well and to transfer the sketch style similar to picture-to-sketch [2, 19].

2 Related Work

Style transfer generates a styled image having similar semantic content as the content image and similar style as the style image. Conventional style transfer is realized by patch-based texture synthesis methods [5, 22] where style is approximated as texture. Given a texture image, patch-based texture synthesis methods

can automatically generate new image with the same texture. However, texture images are quite different from arbitrary style images [22] in duplicated patterns, which limits the functional ability of patch-based texture synthesis method in style transfer. Moreover, control of the texture by varying the patch size (shown in Fig. 2 of [5]) is limited due to the duplicated patterns in the texture image.

The neural style transfer algorithm proposed in [8] is a further development from the work in [7] which pioneers to take advantage of pre-trained CNN on ImageNet [3]. Rather than using texture synthesis methods which are implemented directly on the pixels of raw images, the feature maps of images are used which preserves better semantic information of the image. The algorithm starts with a noise image and finally converges to the styled image by iterative learning. The loss function \mathcal{L} is composed of the content loss $\mathcal{L}_{content}$ and the style loss \mathcal{L}_{style} :

$$\mathcal{L} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style}, \quad (1)$$

$$\mathcal{L}_{content} = \frac{1}{2} (\mathcal{F}_l^{pred} - \mathcal{F}_l^{content})^2, \quad (2)$$

$$\mathcal{L}_{style} = \sum_l e_l \frac{1}{4h_l^2 w_l^2 c_l^2} (G_l^{pred} - G_l^{style})^2, \quad (3)$$

where \mathcal{F}_l^{pred} , $\mathcal{F}_l^{content}$ and \mathcal{F}_l^{style} denote the feature maps of the synthesized styled image, content image and style image separately, \mathcal{F}_l is treated as 2-dimensional data ($\mathcal{F}_l \in \mathcal{R}^{(h_l w_l) \times c_l}$), G_l is the Gram matrix of layer l by $G_l = \mathcal{F}_l^T \times \mathcal{F}_l$, $G_l \in \mathcal{R}^{c_l \times c_l}$, and h_l, w_l, c_l denote the height, width and the channel number of the feature map. Notice that the content loss is measured at the level of feature map, while the style loss is measured at the level of Gram matrix.

Methods were proposed in [9] for spatial control, color control and scale control for neural style transfer. Spatial control is to transfer style of specific regions of the style image via guided feature maps. Given the binary spatial guidance channels for both content and style image, one way to generate the guided feature map is to multiply the guidance channel with the feature map in an element-wise manner while another way is to concatenate the guidance channel with the feature map. Color control is realized by YUV color space and histogram matching methods [10] as post-processing methods. Although both ways are feasible, it is not able to control color to a specified degree, i.e., the control is binary, either transferring all colors of style image or preserving all colors of content image. Moreover, scale control [9] depends on different layers in CNN which represents different abstract levels. Since the number of layers in CNN is finite (19 layers at most for VGG19), the scale of style can only be controlled in finite degrees.

The limitation of control over neural style transfer proposed by pre-processing and post-processing methods in [9] derives from the lack of computational analysis of the artistic style which is the foundation of continuous control for neural

style transfer. Inspired by spatial control in [9] that operations on the feature map could affect the style transferred, we analyze the feature map and decompose the style via projecting feature map into a latent space that is expanded by style basis, such as color, stroke, and so on. Since every point in the latent space can be regarded as a representation of a style and can be decoded back to feature map, style control become continuous and precise. Meanwhile, our work facilitates the mixing or intervention of the style basis from more than one styles so that compound style or new style could be generated, enhancing the diversity of styled images.

3 Methods

We propose to decompose the feature map of the style image into style basis in a latent space, in which it becomes easy to mix or intervene style bases of different styles to generate compound styles or new styles which are then projected back to the feature map space. Such decomposition process enables us to continuously control the composition of the style basis and enhance the diversity of the synthesized styled image. An overview of our method is demonstrated in Fig. 1.

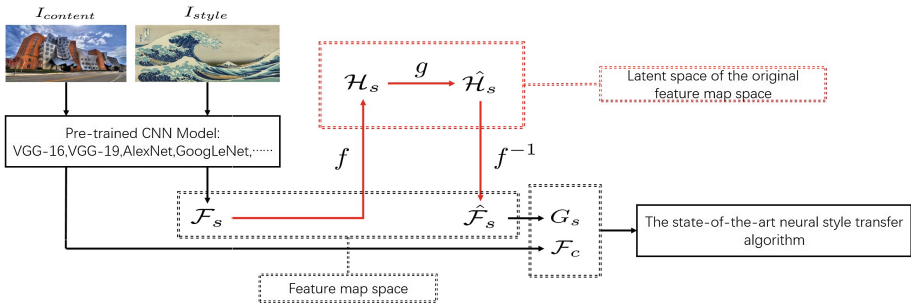


Fig. 1. An overview of our method which is indicated by red part. The red rectangle represents the latent space expanded by the style basis, f denotes computational decomposition of the style feature map \mathcal{F}_s , g denotes mixing or intervention within the latent space. The red part works as a computational module embedded in Gatys’ or other neural style transfer algorithms. (Color figure online)

Given the content image $I_{content}$ and style image I_{style} , we decompose the style by function f from the feature map \mathcal{F}_s of the style image to \mathcal{H}_s in the latent space which is expanded by style basis $\{S_i\}$. We can mix or intervene the style basis via function g which is operated on style basis to generate the desired style coded by $\hat{\mathcal{H}}_s$. Using the inverse function f^{-1} , $\hat{\mathcal{H}}_s$ is projected back to the feature map space to get $\hat{\mathcal{F}}_s$, which replace the original \mathcal{F}_s for style transfer. Our method can serve as embedded module for the state-of-the-art neural style transfer algorithms, as shown in Fig. 1 by red.

It can be noted that the module can be regarded as a general transformation from original style feature map \mathcal{F}_s to new style feature map $\hat{\mathcal{F}}_s$. If we let $\hat{\mathcal{F}}_s = \mathcal{F}_s$, our method degenerates back to traditional neural style transfer [8].

Since the transformation of the feature map is only done on the feature map of the style image, we simply notate \mathcal{F}_s as \mathcal{F} to denote the feature map of the style image and \mathcal{H}_s as \mathcal{H} in the rest of the paper. We notate h and w as the height and width of each channel in the feature map. Next, we introduce two types of decomposition function f and also suggest some control functions g .

3.1 Decomposed by Spectrum Transforms

We adopt 2-dimensional Fast Fourier Transform (FFT) and 2-dimensional Discrete Cosine Transform (DCT) as the decomposition function with details given in Table 1. Both methods are implemented in channel level of \mathcal{F} where each channel is treated as 2-dimensional gray image.

Through the transform by 2-d FFT and 2-d DCT, the style feature map was decomposed as frequencies in the spectrum space where the style is coded by frequency that forms style bases. We will see that some style bases, such as stroke and color, actually correspond to different level of frequencies. With help of decomposition, similar styles are quantified to be close to each other as a cluster in the spectrum space, and it is easy to combine the existing styles to generate compound styles or new styles $\hat{\mathcal{H}}$ by appropriately varying the style codes. $\hat{\mathcal{H}}$ is then projected back to the feature map space via the inverse function of 2-d FFT and 2-d DCT shown in Table 1.

3.2 Decomposed by Latent Variable Models

We consider another type of decomposition by latent variable models, such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA), which decompose the input signal into uncorrelated or independent components. Details are referred to Table 1, where each channel of the feature map \mathcal{F} is vectorized as one input vector.

- **Principal Component Analysis (PCA)**

We implement PCA from the perspective of matrix factorization. The eigenvectors are computed via Singular Value Decomposition (SVD). Then, the style is coded as linear combination of orthogonal eigenvectors, which could be regarded as style bases. By varying the combination of eigenvectors, compound styles or new styles are generated and then projected back to feature map space via the inverse of the matrix of the eigenvectors.

- **Independent Component Analysis (ICA)**

We implement ICA via the fastICA algorithm [11], so that we decompose the style feature map into statistically independent components, which could be regarded as the style bases. Similar to PCA, we could control the combination of independent components to obtain compound styles or new styles, and then project them back to the feature map space.

3.3 Control Function g

The control function g in Fig. 1 defines style operations in the latent space expanded by the decomposed style basis. Instead of operating directly on the feature map space, such operations within the latent space have several advantages. First, after decomposition, style bases are of least redundancy or independent to each other, operations on them are easier to control; Second, the latent space could be made as a low dimensional manifold against noise, by focusing on several key frequencies for the spectrum or principal components in terms of maximum data variation; Third, continuous operations, such as linear mixing, intervention, and interpolation, are possible, and thus the diversity of the output style is enhanced, and even new styles could be sampled from the latent space. Fourth, multiple styles are able to be better mixed and transferred simultaneously.

Table 1. The mathematical details of f and f^{-1}

Method	Decomposition function f	Projection back
2d FFT	$\mathcal{H}(u, v) = \frac{1}{hw} \sum_{x=0}^{h-1} \sum_{y=0}^{w-1} \mathcal{F}(x, y) e^{-2(\frac{ux}{h} + \frac{vy}{w})\pi i}$, $\mathcal{F} \in \mathcal{R}^{h \times w \times c}$	Inverse 2d FFT
2d DCT	$\mathcal{H}(u, v) = c(u)c(v) \sum_{x=0}^{h-1} \sum_{y=0}^{w-1} \mathcal{F}(x, y) \cos[\frac{(x+0.5)\pi}{h} u] \cos[\frac{(y+0.5)\pi}{w} v]$ $\mathcal{F} \in \mathcal{R}^{h \times w \times c}$, $c(u) = \sqrt{\frac{1}{N}}$, $u = 0$ and $c(u) = \sqrt{\frac{2}{N}}$, $u \neq 0$	Inverse 2d DCT
PCA	$\mathcal{F} = UDV^T$, $U = [v_1, \dots, v_{hw}]$, $\mathcal{H} = U \times \mathcal{F}$ $\mathcal{F} \in \mathcal{R}^{(hw) \times c}$, $U \in \mathcal{R}^{(hw) \times (hw)}$	$\hat{\mathcal{F}} = U^T \times \hat{\mathcal{H}}$
ICA	$[S, A] = \text{fastICA}(\mathcal{F})$, $\mathcal{H} = S$, $\mathcal{F} \in \mathcal{R}^{(hw) \times c}$ $S \in \mathcal{R}^{c \times (hw)}$, $A \in \mathcal{R}^{c \times c}$	$\hat{\mathcal{F}} = (A \times \hat{\mathcal{H}})^T$

Let $S_i^{(n)}$, $i \in \mathbb{Z}$ denote the i -th style basis of n -th style image. Notate $\{S_i^{(n)} | i \in I\}$, $I \subset \mathbb{Z}$ as $S_I^{(n)}$.

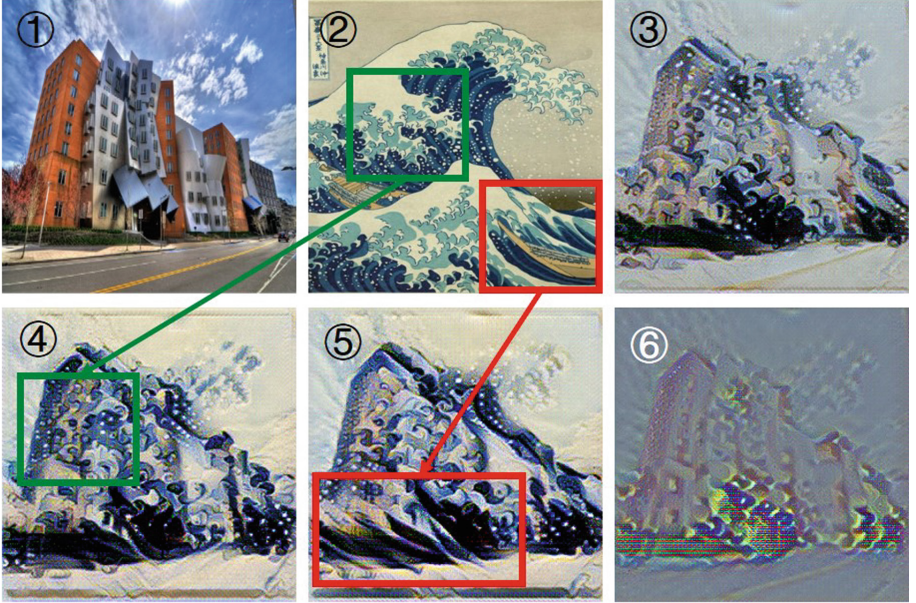
- **Single style basis:** Project the latent space on one of the style basis S_j . That is $S_i = 0$ if $i \neq j$
- **Intervention:** Reduce or amplify the effect of one style basis S_j by multiplying I while keeping other style bases unchanged. That is $S_i = I * S_i$ if $i = j$
- **Mixing:** Combine the style bases of n styles. That is $S = \{S_I^{(1)}, S_J^{(2)}, \dots, S_K^{(n)}\}$

4 Experiments

We demonstrate the performance of our method using the fast neural style transfer algorithm [6, 13, 21]. We take the feature map ‘relu4.1’ from pre-trained VGG-19 model [18] as input to our style decomposition method because we try every single activation layer in VGG-19 and find that ‘relu4.1’ is suitable for style transfer.

4.1 Inferiority of Feature Map

Here, we demonstrate that it is not suitable for the style control function g to be applied on the feature map space directly because feature map space is possibly formed by a complicated mixture of style bases. To check whether the basis of feature map \mathcal{F} can form the style bases, we experimented on the channels of \mathcal{F} and the pixels of \mathcal{F} .



(a)

Fig. 2. (a) ① content image (Stata Center); ② style image (“The Great Wave off Kanagawa” by Katsushika Hokusai); ③ styled image by traditional neural style transfer; ④⑤⑥ are the results of implementing control function directly on the feature map \mathcal{F} . Specifically, we amplify some pixels of \mathcal{F} which generate ④⑤ and preserve a subset of channels of \mathcal{F} which generate ⑥. (Color figure online)

Channels of \mathcal{F} . Assume styles are encoded in space \mathcal{H} which is expanded by style basis $\{S_1, S_2, \dots, S_n\}$. A good formulation of \mathcal{H} may imply that if two styles are intuitively similar from certain aspects, there should be at least one style basis S_i such that the projections of the two styles onto S_i are close in Euclidean distance. Based on the above assumption, we generate the subset \mathcal{C} of channels of \mathcal{F} that could possibly represent color basis with semi-supervised method using style images in Fig. 13(a–c). It can be noticed that Chinese paintings and pen sketches (Fig. 13(a, c)) share the same color style while oil painting

(Fig. 13(b)) has an exclusive one. We iteratively find the largest channel set C_{max} (384 channels included) whose clustering result out of K-means [16] conforms to the following **clustering standard for color basis**:

- No cluster contains both oil painting and Chinese painting or pen sketch.
- One cluster contains only one or two points, since K-means is not adaptive to the cluster number and the cluster number is set as 3.

However, if we only use C_{max} to transfer style, the styled image (Fig. 2(a)⑥) isn't well stylized and doesn't indicate any color style of the style image (Fig. 2②), which probably indicates that the channels of \mathcal{F} are not suitable to form independent style basis. However, by proper decomposition functions f introduced in Table 1, a good formulation of latent space \mathcal{H} is feasible to reach the above clustering standard and to generate reasonably styled images.

Pixels of \mathcal{F} . We give intervention $I = 2.0$ to certain region of each channel of \mathcal{F} to see if any intuitive style basis is amplified. The styled images are shown in Fig. 2(a)④⑤. Rectangles in style image (Fig. 2(a)②) are the intervened regions correspondingly. Compared to the styled image using [8] (Fig. 2(a)③), when small waves in style image is intervened, the effect of small blue circles in the styled image are amplified (green rectangle) and when large waves in style image is intervened, the effect of long blue curves in the styled image are amplified (red rectangle). Actually, implementing control function g on the pixels of the channels of \mathcal{F} is quite similar to the methods proposed for spatial control of neural style transfer [9] which controls style transfer via a spatially guided feature map defined by a binary or real-valued mask on a region of the feature map. Yet it fails to computationally decompose the style basis.

4.2 Transfer by Single Style Basis

To check whether \mathcal{H} is composed of style bases, we transfer style with single style basis preserved. We conduct experiments on \mathcal{H} formulated by different decomposition functions, including FFT, DCT, PCA as well as ICA, with details mentioned in Sect. 3. The results are shown in Fig. 3. As is shown in Fig. 3(c), (d), The DC component only represents the color of style while the rest frequency components represent the wave-like stroke, which indicates that FFT is feasible for style decomposition. The result of DCT is quite similar to that of FFT, with DC component representing color and the rest representing stroke.

Besides, we analyze the spectrum space via Isomap [20], which can analytically demonstrate the effectiveness and robustness of spectrum based methods. Color (low frequency) and stroke (high frequency) forms X-axis and Y-axis of the 2-dimensional plane respectively where every style is encoded as a point. We experiment on 3 artistic styles, including Chinese painting, oil painting and pen sketch, and each style contains 10 pictures which is shown in Fig. 13(a–c). Chinese paintings and pen sketches share similar color style which is sharply

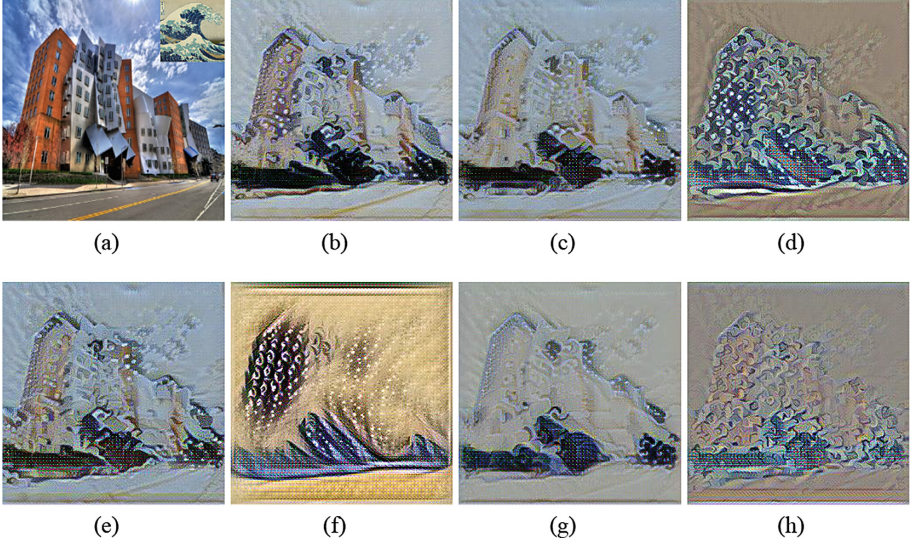


Fig. 3. (a) the original content and style images; (b) styled image by traditional neural style transfer; (c–h) results of preserving one style basis by different methods. Specifically, (c–d) FFT; (e–f) PCA; (g–h) ICA where (c, e, g) aim to transfer the color of style image while (d, f, h) aim to transfer the stroke of style image.

distinguished with oil paintings’ while the stroke of three artistic styles are quite different from each other, which conforms to the result shown as Fig. 13(d).

Via PCA, the first principal component (Fig. 3(e)) fails to separate color and stroke, while the rest components (Fig. 3(f)) fail to represent any style basis, which indicates PCA is not suitable for style decomposition (Fig. 4).

The results of ICA (Fig. 3(g), (h)) are as good as the results of FFT but show significant differences. The color basis (Fig. 3(g)) is more murky than Fig. 3(c) while the stroke basis (Fig. 3(h)) retains the profile of curves with less stroke color preserved compared to Fig. 3(d). The stroke basis consists of $S_{arg_i}, i \in [0, n - 1] \cup [c - n, c - 1]$ while the rest forms the color basis. $arg \in \mathcal{R}^c$ is the ascent order of $A^{sum} \in \mathcal{R}^c$ where A_i^{sum} is the sum of i th column of mixing matrix A .

4.3 Transfer by Intervention

We give intervention to the stroke basis via control function g to demonstrate the controllable diversified styles and distinguish the difference in stroke basis between spectrum based methods and ICA.

As is shown in Fig. 5, the strokes of ‘wave’ are curves with light and dark blue while the strokes of ‘aMuse’ are black bold lines and coarse powder-like dots.

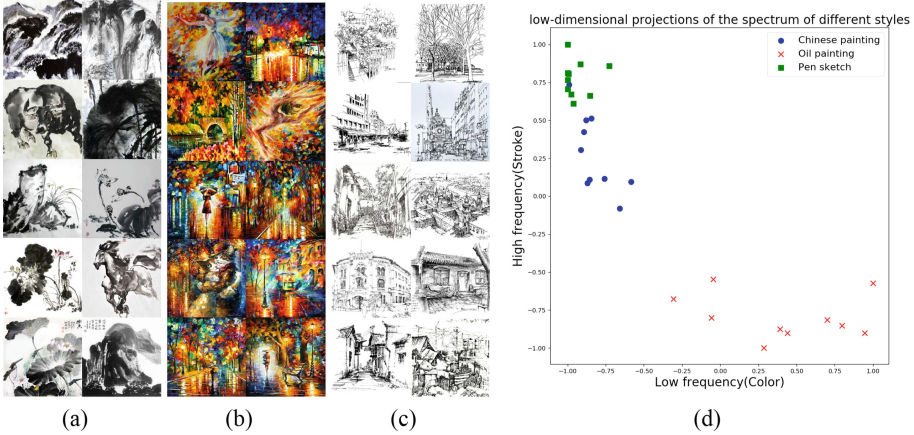


Fig. 4. (a) Chinese paintings; (b) Oil paintings (by Leonid Afremov); (c) Pen sketches; (d) low-dimensional projections of the spectrum of style(a-c) via Isomap; (e) low-dimensional projections of the spectrum of large scale of style images via Isomap. The size of each image shown above does not indicate any other information, but is set to prevent the overlap of the images only.

Intervention using spectrum based methods affects both color and outline of the strokes while intervention using ICA only influences stroke outline but not its color.

4.4 Transfer by Mixing

Current style mixing method, interpolation, cannot mix the style bases of different styles because styles are integrally mixed however interpolation weights are modified (Fig. 6(g-i)), which limits the diversity of style mixing. Based on the success spectrum based methods and ICA in style decomposition, we experiment to mix the stroke of ‘wave’ with the color of ‘aMuse’ to check whether such newly compound artistic style can be transferred to the styled image.

Specifically, for ICA, we not only need to replace the color basis of ‘wave’ with that of ‘aMuse’ but also should replace the rows of mixing matrix A corresponding to the exchanged signals. Both spectrum based methods (Fig. 6(d-f)) and ICA (Fig. 6(j-l)) work well in mixing style bases of different styles. Moreover, we can intervene the style basis when mixing, which further enhances the diversity of style mixing.

4.5 Sketch Style Transfer

The picture-to-sketch problem challenges whether computer can understand and represent the concept of objects both abstractly and semantically. The proposed controllable neural style transfer tackles the obstacle in current State-of-the-art

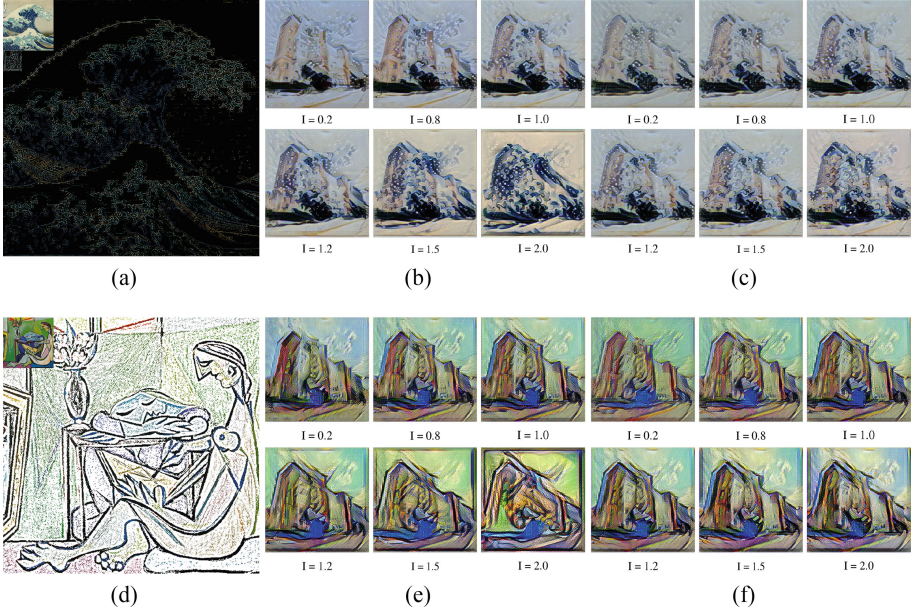


Fig. 5. (a, d) the stroke of style image ‘wave’ and ‘aMuse’; (b, c, e, f) the results with different intervention I to the stroke basis. Specifically, (b, e) use spectrum based method and (c, f) use ICA. (Color figure online)

methods [2, 19] which is caused by inconsistency in sketch style because diversified styles can in turn increase the style diversity of output images. Moreover, as is shown in Fig. 7, our method can control the abstract level by reserving major semantic details and minor ones automatically. Our method does not require vector sketch dataset. As a result, we cannot generate sketches in a stroke-drawing way [2, 19].

4.6 Chinese Painting Style Transfer

Chinese painting is an exclusive artistic style having much less color than Western painting and represents the artistic conception by strokes. As is shown in Fig. 8, with effective controls over stroke via our methods, the Chinese painting styled image can be either misty-like known as freehand-brush or meticulous representation known as fine-brush.

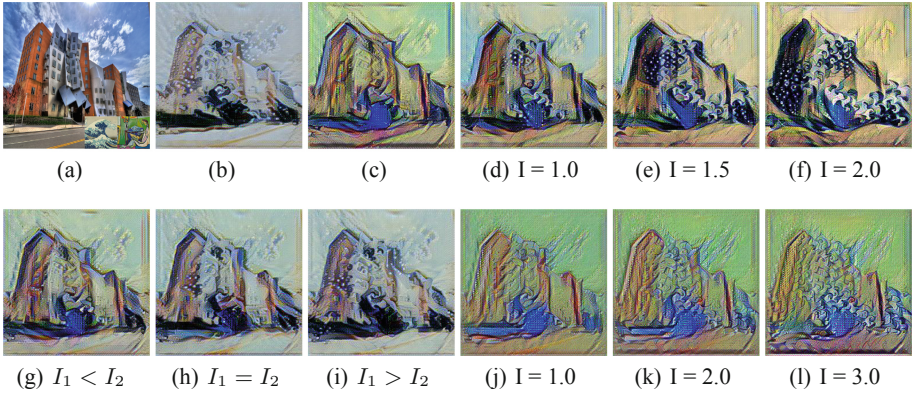


Fig. 6. (b), (c) styled image of single style; (g–i) interpolation mixing where I_1 and I_2 are the weights of ‘wave’ and ‘aMuse’; (d–f, j–l) results of mixing the color of ‘aMuse’ and the stroke of ‘wave’ where I is the intervention to the stroke of ‘wave’. Specifically, (d–f) use FFT; (j–l) use ICA.

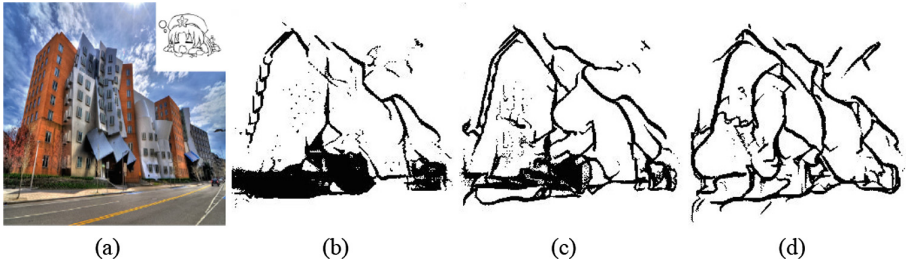


Fig. 7. Picture-to-sketch using style transfer and binarization. (a) content image and style image; (b–d) styled images. From (b) to (d), the number of stroke increases as more details of the content image are restored.

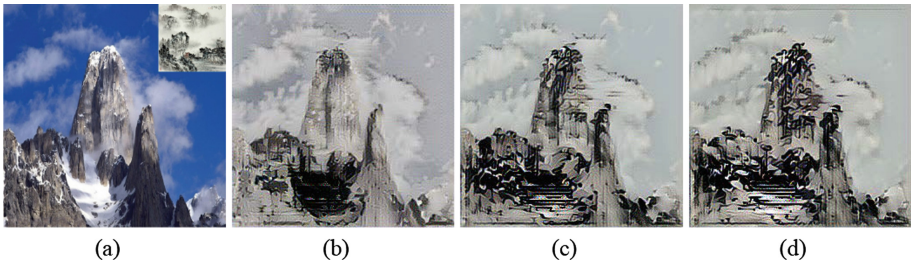


Fig. 8. Neural style transfer of Chinese painting with stroke controlled. (a) content image and style image (by Zaixin Miao); (b–d) styled images. From (b) to (d), the strokes are getting more detailed which gradually turns freehand style into finebrush style.

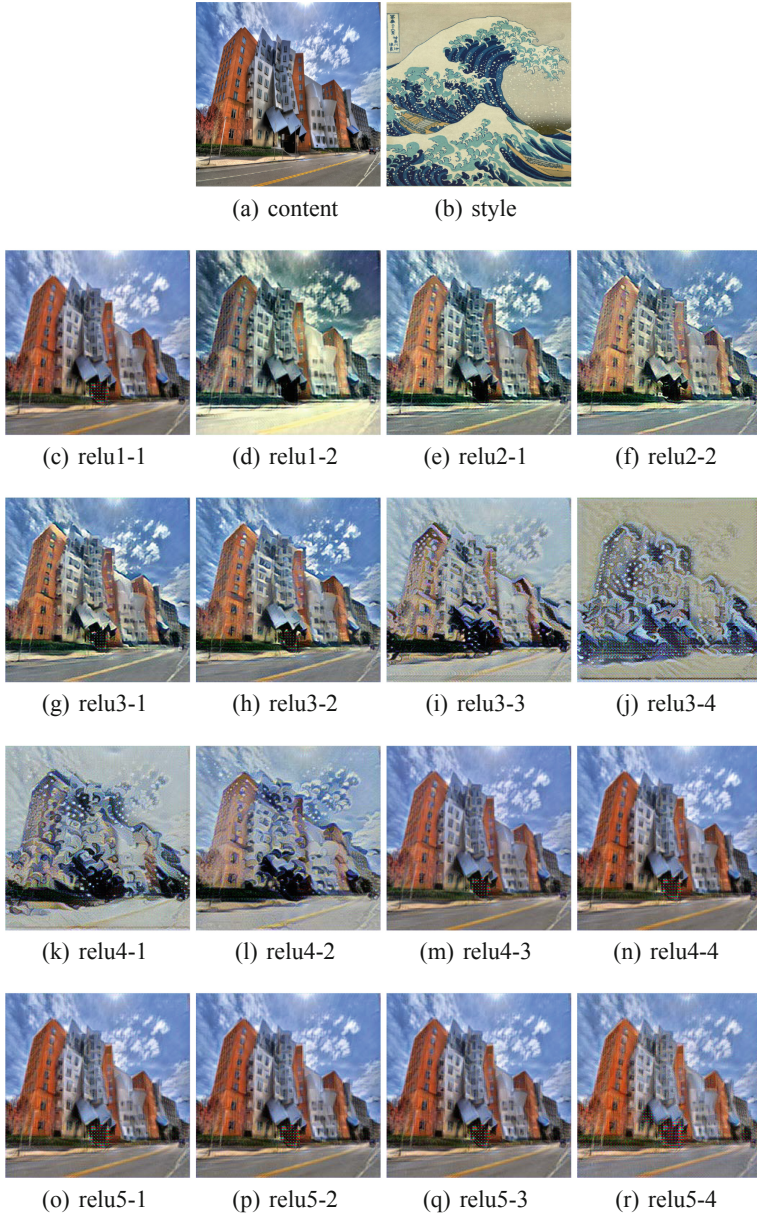


Fig. 9. Styled images using [8] with same epochs using every single activation layer from the pre-trained VGG19.

5 Conclusion

Artistic styles are made of basic elements, each with distinct characteristics and functionality. Developing such a style decomposition method facilitate the quantitative control of the styles in one or more images to be transferred to another natural image, while still keeping the basic content of natural image. In this paper, we proposed a novel computational decomposition method, and demonstrated its strengths via extensive experiments. The method could serve as a computational module embedded in those neural style transfer algorithms. We implemented the decomposition function by spectrum transform or latent variable models, and thus it enabled us to computationally and continuously control the styles by linear mixing or intervention. Experiments showed that our method enhanced the flexibility of style mixing and the diversity of stylization. Moreover, our method could be applied in picture-to-sketch problems by transferring the sketch style, and it captures the key feature and facilitates the stylization of the Chinese painting style.

Acknowledgement. This work was supported by the Zhi-Yuan Chair Professorship Start-up Grant (WF220103010), and Startup Fund (WF220403029) for Youngman Research, from Shanghai Jiao Tong University.

Appendix 1: The Stylization Effect of Every Activation Layer in VGG19

Since different layers in VGG19 [18] represent different abstract levels, we experiment the stylization effect of every activation layer on a couple of different style images as are shown in Figs. 9 and 10. From our experiments, it can be noticed that not every layer is effective in style transfer and among those that work, shallow layers only transfer the coarse scale of style (color) while deep layers can transfer both the coarse scale (color) and detailed scale (stroke) of style, which conforms to the result of scale control in [9]. Since ‘relu4_1’ performs the best in style transfer after same learning epochs, we determine to study the feature map of ‘relu4_1’ in our research.

We further visualize each channel of the feature map of the style image using t-SNE [4], as are shown in Figs. 11 and 12 where the similarity of both the results is quite interesting. However, we could not explain those specific patterns shown in the visualization results yet. What’s more, the relationship between the similarity of the visualization results with the similarity of the stylization effects of different VGG layers remains further study as well.

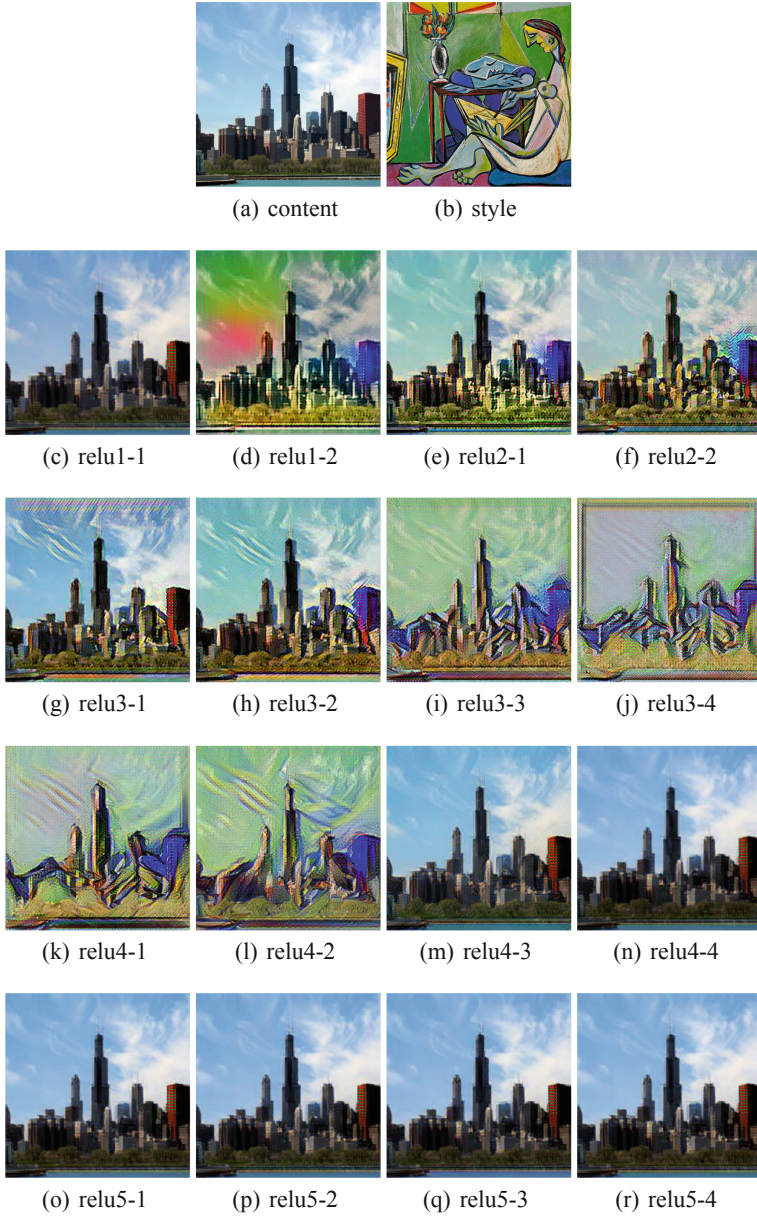


Fig. 10. Styled images using [8] with same epochs using every single activation layer from the pre-trained VGG19.

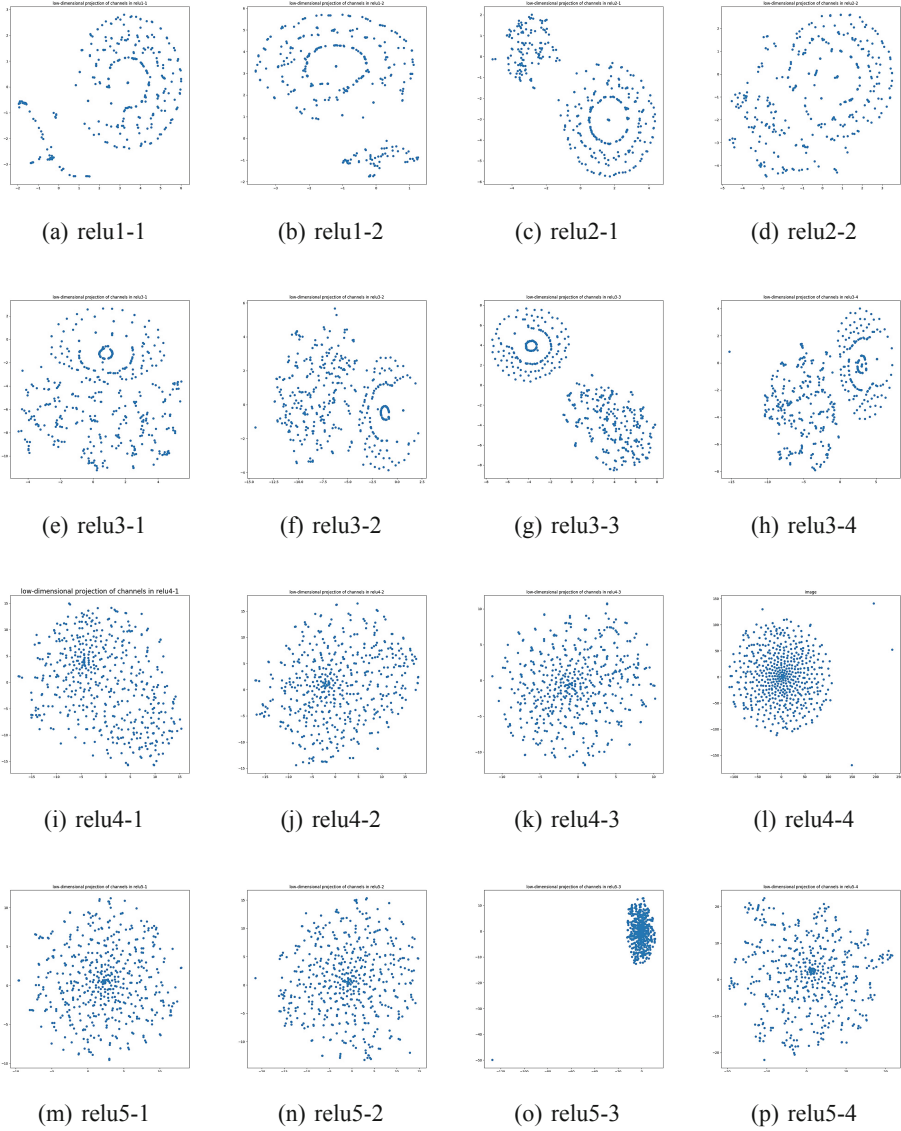


Fig. 11. Low dimensional projection of all channels of every single layer of the style image (Fig. 9(b)) via t-SNE [4].

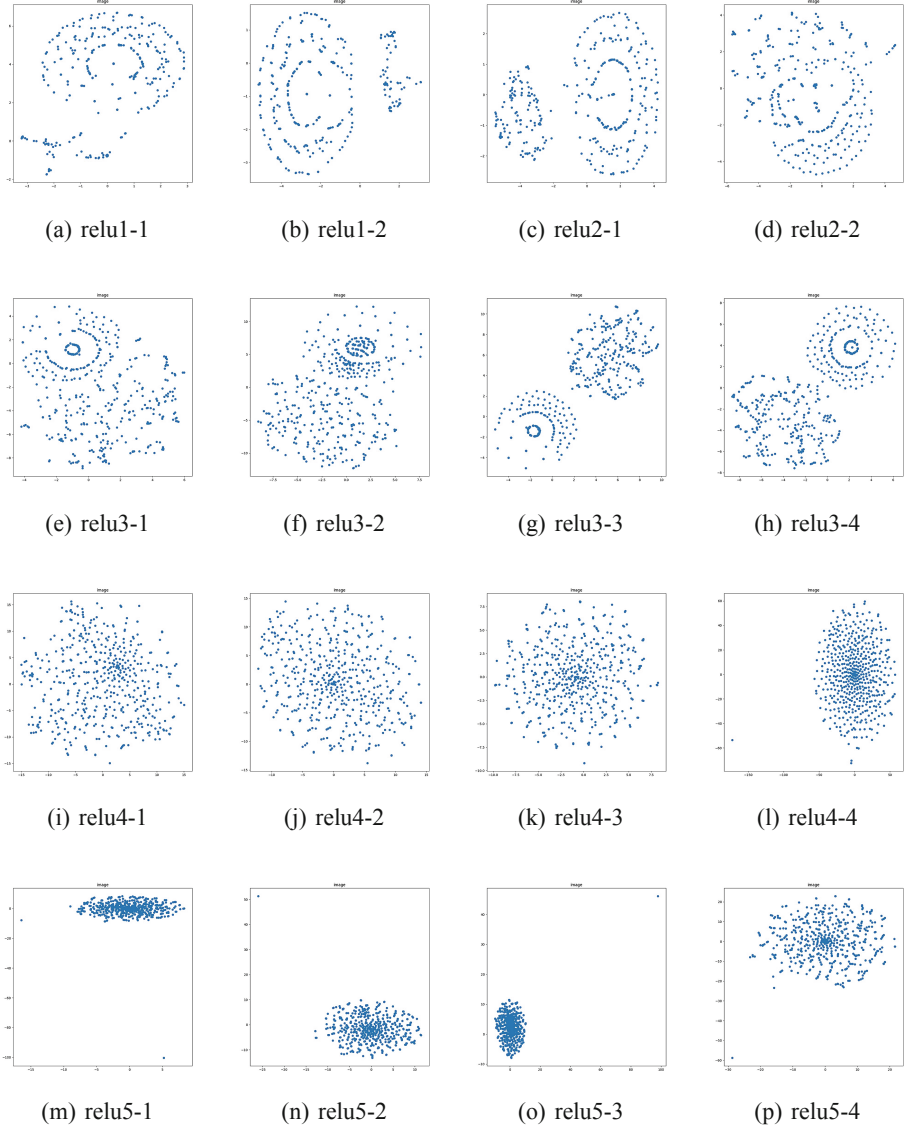


Fig. 12. Low dimensional projection of all channels of every single layer of style image (Fig. 10(b)) via t-SNE [4].

Appendix 2: The Manifold of Spectrum Based Methods

We analyze the spectrum space by projecting the style bases via Isomap [20] into low dimensional space where the X-axis represents the color basis and the Y-axis represents the stroke basis, which can analytically demonstrate the effectiveness and robustness of spectrum based methods. Three artistic styles are experimented (shown in Fig. 13(a-c)). Chinese paintings and pen sketches share

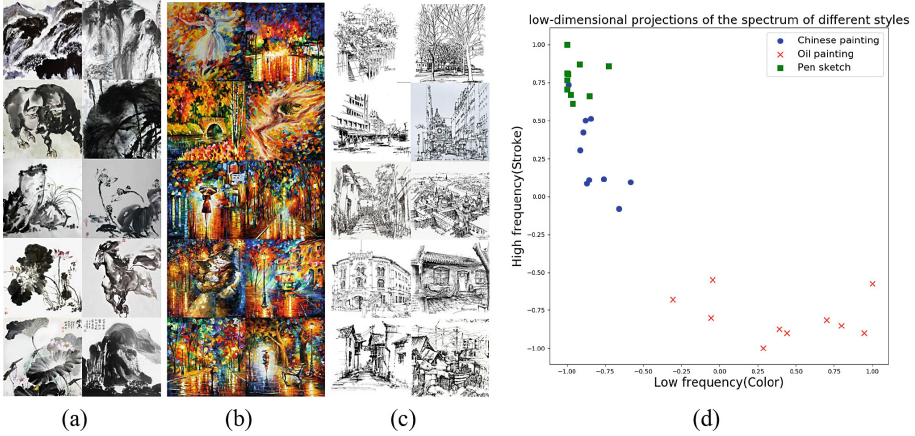


Fig. 13. (a) Chinese paintings; (b) Oil paintings (by Leonid Afremov); (c) Pen sketches; (d) low-dimensional projections of the spectrum of style (a-c) via Isomap [20].

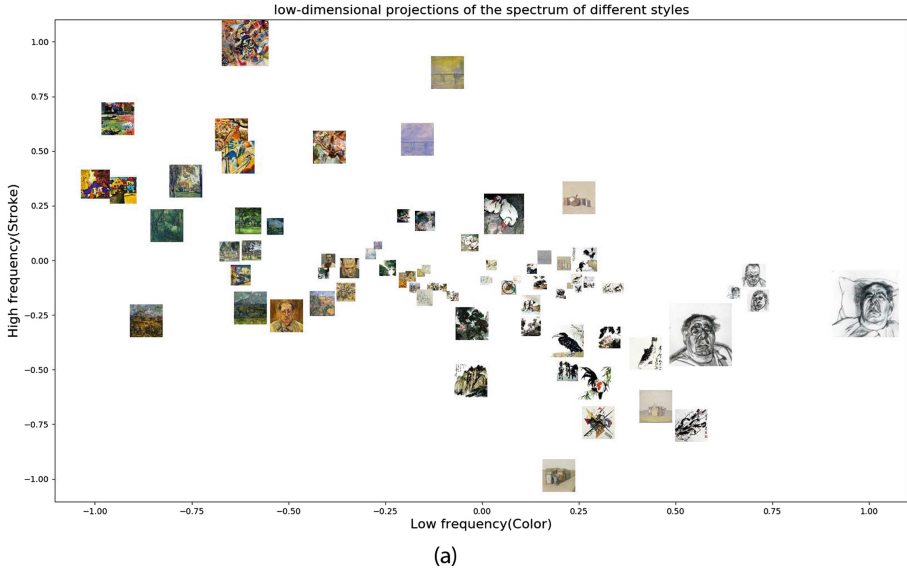


Fig. 14. Low-dimensional projections of the spectrum of large scale of style images via Isomap [20]. The size of each image shown above does not indicate any other information, but is set to prevent the overlap of the images only.

similar color style which is sharply distinguished with oil paintings' while the stroke of three artistic styles are quite different from each other. Thus, as is shown in Fig. 13(d), Chinese paintings and pen sketches are close to each other and both stay away from oil paintings in X-axis which represents color while three styles are respectively separable in Y-axis which represents stroke, which completely satisfies our analysis of the three artistic styles.

When we apply the same method to large scale of style images (Fig. 14), X-axis clearly represents the linear transition from dull-colored to rich-colored. However, we fail to conclude any notable linear transition for Y-axis from the 2-dimensional visualization probably because it is hard to describe the style of stroke (boldface, length, curvity, etc.) using only one dimension.



Fig. 15. The styled image with stroke basis intervened using spectrum based methods. The left most row shows the style images (From top to bottom: The Great Wave off Kanagawa - Katsushika Hokusai; Composition - Alberto Magnelli; Dancer - Ernst Ludwig Kirchner; Pistachio Tree in the Courtyard of the Chateau Noir - Paul Cezanne; Portrait - Lucian Freud). From left to right of each row, the effect of stroke is increasingly amplified.

Appendix 3: Stroke Intervention

We demonstrate more styled images with stroke basis intervened using spectrum based method (Figs. 15 and 16) and ICA (Figs. 17 and 18) respectively.



Fig. 16. The styled image with stroke basis intervened using spectrum based methods. The left most row shows the style images (From top to bottom: A Muse (La Muse) - Pablo Picasso; Number 4 (Gray and Red) - Jackson Pollock; Shipwreck - J.M.W. Turner; Natura Morta - Giorgio Morandi; The Scream - Edvard Munch). From left to right of each row, the effect of stroke is increasingly amplified. (Color figure online)



Fig. 17. The styled image with stroke basis intervened using ICA. The left most row shows the style images (From top to bottom: The Great Wave off Kanagawa - Katsushika Hokusai; Composition - Alberto Magnelli; Dancer - Ernst Ludwig Kirchner; Pistachio Tree in the Courtyard of the Chateau Noir - Paul Cezanne; Potrait - Lucian Freud). From left to right of each row, the effect of stroke is increasingly amplified.

Appendix 4: Style Mixing

We demonstrate more styled images transferred with compound style generated by mixing the color basis and stroke basis of two different styles. The results of both spectrum based method and ICA method are shown in Fig. 19 with comparison with traditional mixing method - interpolation.



Fig. 18. The styled image with stroke basis intervened using ICA. The left most row shows the style images (From top to bottom: A Muse (La Muse) - Pablo Picasso; Number 4 (Gray and Red) - Jackson Pollock; Shipwreck - J.M.W. Turner; Natura Morta - Giorgio Morandi; The Scream - Edvard Munch). From left to right of each row, the effect of stroke is increasingly amplified. (Color figure online)

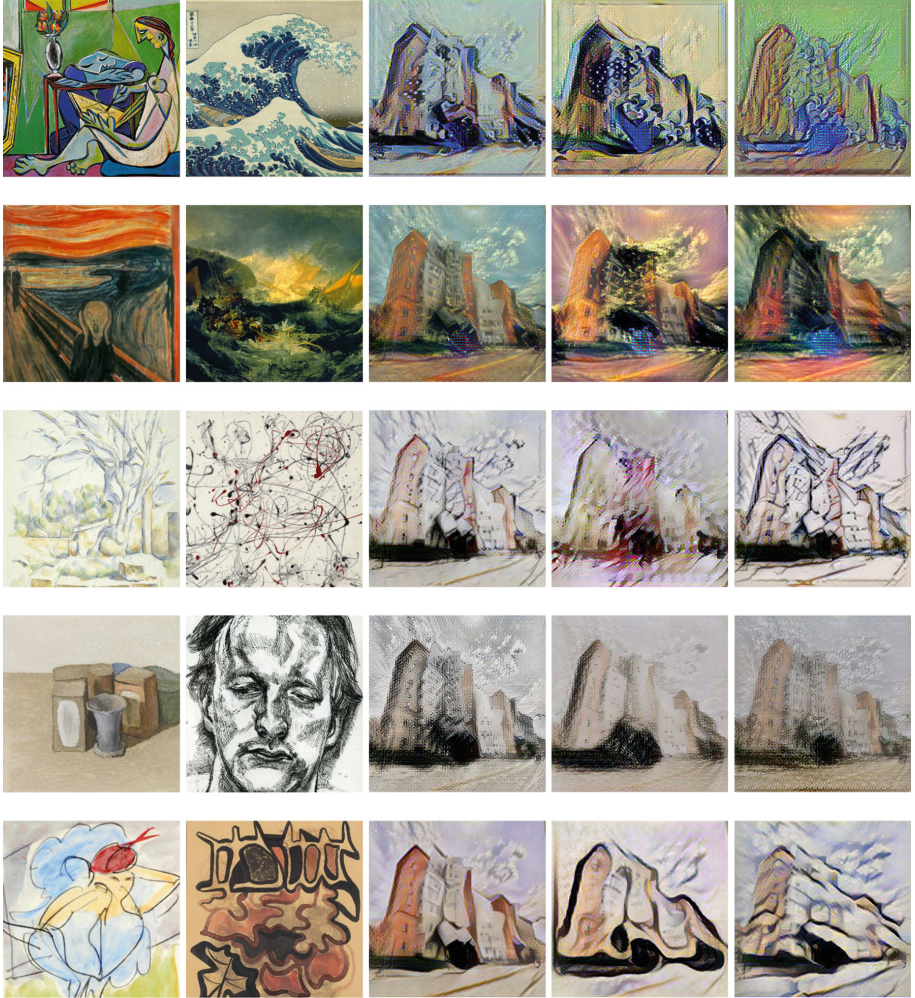


Fig. 19. The left two columns are the style images used for mixing. Specifically, we mix the color of the most left one with the stroke of the second left one. The third left column shows the styled images with traditional interpolation method. The second right column shows the styled images using spectrum mixing method. The most right column shows the styled images using ICA mixing method.

References

1. Champanard, A.J.: Semantic style transfer and turning two-bit doodles into fine artworks. arXiv preprint [arXiv:1603.01768](https://arxiv.org/abs/1603.01768) (2016)
2. Chen, Y., Tu, S., Yi, Y., Xu, L.: Sketch-pix2seq: a model to generate sketches of multiple categories. CoRR abs/1709.04121 (2017)

3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
4. Der Maaten, L.V., Hinton, G.E.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008)
5. Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1033–1038. IEEE (1999)
6. Engstrom, L.: Fast style transfer (2016). <https://github.com/lengstrom/fast-style-transfer/>
7. Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 28, pp. 262–270. Curran Associates, Inc. (2015)
8. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015)
9. Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E.: Controlling perceptual factors in neural style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3985–3993 (2017)
10. Hertzmann, A.P.: Algorithms for rendering in artistic styles. Ph.D. thesis, New York University, Graduate School of Arts and Science (2001)
11. Hyvarinen, A.: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10**(3), 626–634 (1999)
12. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: a review. *IEEE Trans. Vis. Comput. Graph.* (2019)
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
14. Li, C., Wand, M.: Combining Markov random fields and convolutional neural networks for image synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2479–2486 (2016)
15. Li, C., Wand, M.: Precomputed real-time texture synthesis with Markovian generative adversarial networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9907, pp. 702–716. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_43
16. Lloyd, S.P.: Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**(2), 129–137 (1982)
17. Rosin, P., Collomosse, J.: *Image and Video-Based Artistic Stylisation*, vol. 42. Springer, London (2012). <https://doi.org/10.1007/978-1-4471-4519-6>
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
19. Song, J., Pang, K., Song, Y.Z., Xiang, T., Hospedales, T.M.: Learning to sketch with shortcut cycle consistency. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 801–810 (2018)
20. Tenenbaum, J.B., De Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**(5500), 2319–2323 (2000)
21. Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.: Texture networks: feed-forward synthesis of textures and stylized images. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, ICML 2016, vol. 48, pp. 1349–1357 (2016)

22. Wei, L.Y., Levoy, M.: Fast texture synthesis using tree-structured vector quantization. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2000, pp. 479–488. ACM Press/Addison-Wesley Publishing Co., New York (2000)
23. Yin, R.: Content aware neural style transfer. CoRR abs/1601.04568 (2016). <http://arxiv.org/abs/1601.04568>